

# Data Compression - Systematisation

## ---- Glossary ----

by T.Strutz

### **back to Start**

alphabet	... entirety of all (different) symbols in a certain context
bitrate	... average number of bits per symbols for storage or transmission
code	... entirety of code words of a alphabet
code word	... string of bits
code length	... number of bits forming a code word
code value	... value of a bit string, if read as binary number
compression ratio	... quotient of storage amount before and after the compression
lossless compression	... compression without changes of digital information, data reduction methods are not allowed
lossy compression	... compression comes with changes of digital information due to application of data reduction methods
reconstruction error	... difference between original and reconstructed signal

## What is Irrelevancy?

Here a definition:

Irrelevancy comprises all those parts of the signal content, which either can not be sensed by the signal recipient or are unimportant for the recipient.

Thus, if we want to describe irrelevancy, we have always to consider, who the signal receives. Here an example:

Imagine we have an ASCII text containing formatted source code in the programming language C, like this:

```
/*
 * This is a Hello-World demo program.
 */
#include <stdio.h>

int main( int argc, char *argv[])
{
    if (argc > 1)
    {
        printf( "Hello World!\n");
    }
}
```

```
    exit( 0);
}
```

Now lets imagine any compiler program is the final recipient. What is irrelevant? All tabs, new-lines, and most of the spaces as well as all comments! A good data reduction would produce following:

```
#include<stdio.h>int    main(int    argc,char*argv[]){if(argc>1){printf
("Hello World!\n");}exit(0);}
```

However, what have we done, should a human programmer want to modify the program? We have removed a lot of important information! Ok, we could reconstruct troublesomely the formatting by reinserting white-characters, however, the comments are lost for ever.

**back to Start**

Strutz / 20.02.2003

